DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

ANALYSIS OF INTELLIGENT BUS STOP DATA TRAFFIC BY USING KENTKART DATABASE

by
Deniz KARABAŞ

February, 2016
İZMİR

ANALYSIS OF INTELLIGENT BUS STOP DATA TRAFFIC BY USING KENTKART DATABASE

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Master of Science
In Computer Engineering Program

by
Deniz KARABAŞ

February, 2016
İZMİR

M.Sc THESIS EXAMINATION RESULT FORM

We have read the thesis entitled "**ANALYSIS OF INTELLIGENT BUS STOP DATA TRAFFIC BY USING KENTKART DATABASE**" completed by **DENİZ KARABAŞ** under supervision of **ASST. PROF. DR SEMİH UTKU** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

....

Asst. Prof. Dr. Semih UTKU

Supervisor

....

Asst. Prof. Dr. Canan Atay   Assoc. Prof. Dr. Mehmet S. Ünlütürk

(Jury Member)   (Jury Member)

Prof. Dr. Ayşe OKUR

Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENTS

I would like to thank to my supervisor, Assistant Professor Dr. Semih UTKU, for his support, supervision and useful suggestions throughout this study.

This study is developed as a part of the project (Project Number 3140852) supported by KENTKART Company and TUBITAK (Türkiye Bilimsel ve Teknik Araştırma Kurumu) to whom I wish to thank for their support.

I would like to dedicate this work to my husband Sefa KARABAŞ for his support while writing the thesis.

I also would like to thank to my friend Atilla SUNCAK for his support.

I would like to offer my special thanks to my family for their support, patience and help.

Deniz KARABAŞ

ANALYSIS OF INTELLIGENT BUS STOP DATA TRAFFIC BY USING KENTKART DATABASE

ABSTRACT

Increases in traffic density over years have been causing waiting time on roads to increase and longer travel time. Therefore, smart solutions have been arisen in order to inform passengers and provide them more efficient and reliable public transportation. The important part of the transportation system is the Passenger Information System.

As a lot of people use *buses* in order to go to work, school and another place every day, they need the passenger information system quite a lot. Most of the complaints of passengers are about spending time at the bus stops for waiting buses. The main goal of this study is to minimize the waiting time of passengers at bus stops by predicting the arrival time of buses.

There have been several studies performed on predicting the arrival time of buses at bus stops. Some of the methods which have been widely used to provide data are sensors which define the number of vehicles that pass by fixed points and camera-based sensors. Storing and analyzing data in those ways are extremely expensive. Another method is predicting how much time later a bus will arrive at the following bus stop by using GPS data in vehicles. In the study, certain pieces of information such as buses' location, speed, the last bus stop they pass by, arrival time at the last bus stop they pass by are transferred to the system owing to GPS receivers in buses. In order to estimate the arrival time of buses to the bus stops, Kalman filtering and average travel time algorithms have been benefited. The study has been performed by means of GPS data gathered from the buses using Kentkart system in Manisa.

Keywords: Intelligent transportation system, global positioning system, Kalman filter, average time algorithm.

KENTKART VERİTABANI KULLANILARAK AKILLI DURAK VERİ TRAFİĞİNİN ANALİZİ

ÖZ

Yıldan yıla trafik hacminde oluşan artış, yollarda bekleme sürelerini artırmakta, seyahat sürelerinin uzamasına neden olmaktadır. Bu noktada yolcuları bilgilendirmek ve onların daha etkin ve güvenilir bir şekilde seyahat edebilmeleri için akılcı çözümler ortaya çıkmıştır. Ulaşım sistemlerinin önemli bir parçasıda yolcu bilgilendirme sistemidir.

Günümüzde yolcu bilgilendirme sistemine en çok ihtiyaç duyan toplu ulaşım araçları otobüslerdir. Çünkü her gün pek çok insan işe, okula ya da herhangi bir yere gitmek için otobüsü kullanmaktadır.Yolcuların en çok şikayet ettikleri durumlardan biri de belirsiz bir zaman süresince otobüs bekleyerek zaman kaybetmektir. Bu çalışmanın amacı, otobüslerin duraklara varış sürelerini tahminleyerek kullanıcıların duraklarda bekleme sürelerini en az hale getirmektir.

Duraklara varış sürelerinin tahminlemesi ile ilgili pek çok çalışma yapılmıştır. Bu çalışmalarda en çok kullanılan veri sağlayıcı yöntemlerinden bazıları; sabit noktalardan geçen araç sayılarını belirleyen algılayıcılar; iz tabanlı algılayıcılar; kamera tabanlı algılayıcılardır. Bu yöntemler ile veriyi saklamak ve analiz etmek oldukça pahalıdır. Diğer bir yöntemde araçlardaki gps verisinin kullanılarak bir sonraki durağa ne kadar sürede gelebileceğinin tahminlenmesidir. Bu çalışmada, otobüslerin içindeki gps alıcıları sayesinde otobüslerin konum, hız, son geçtiği durak bilgisi ve son geçtiği durağa varış süresi gibi bilgilerin sisteme aktarılması sağlanmıştır. Duraklara varış sürelerinin tahminlenmesi için Kalman filtresi ve ortalama süre algoritması kullanılmıştır. Bu çalışma, Kentkart sisteminde yer alan Manisa ilindeki otobüslerin gps verileri kullanılarak gerçekleştirilmiştir.

Anahtar Kelimeler: Akıllı taşıma sistemleri, küresel konumlama sistemi (GPS), Kalman filtresi, ortalama süre algoritması.

CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER ONE
INTRODUCTION

1.1 Overview

Increase in traffic volume has been posing a major threat to the quality of public transportation. This situation causes traffic congestion, which leads to time loss and increase in arrival time to the destination. A good public transportation system has to be highly qualified, reliable and accessible.

In Turkey, buses, trains and ferries are mostly used public transportation services. It is easier to estimate the arrival time of trains and ferries than estimating the arrival time of buses at bus stops, because trains and ferries have stationary speed and travel time of them is usually stable. Estimating the arrival time of a bus to the bus stop is important for passengers, because if a passenger misses a bus, it will take a long time to wait for the next one or find an alternative way of transportation. Passengers aren't concerned about time that they spend while waiting at the bus stop as long as they know the arrival time of alternative buses at the bus stop. However, if the passenger doesn't know the arrival time of the next bus, then he/she becomes concerned about delays or the level of traffic congestion.

The most important component of an advanced and reliable public transportation is the *Intelligent Transportation System* (ITS) which consists of several methods, techniques and methodologies for passenger information system. Those methods provide reducing time, financial loss and informing to passengers. Moreover, many methods are used for estimating the arrival time of a bus to the bus stop which increases reliability and accuracy of the travel time estimation. As it is shown in Figure 1.1, there are several factors which affect the travel time estimation such as road geometry, weather conditions, time of day, day of week, population of the city and etc. If historical data are processed, and then all those factors should be considered in order to obtain more reliable results, however this study performs real time data therefore it is not essential to use all the factors. Estimation of travel time at the bus stops is calculated according to the last GPS data of the buses which give the road status on the current time; therefore the results will be more reliable and accurate.

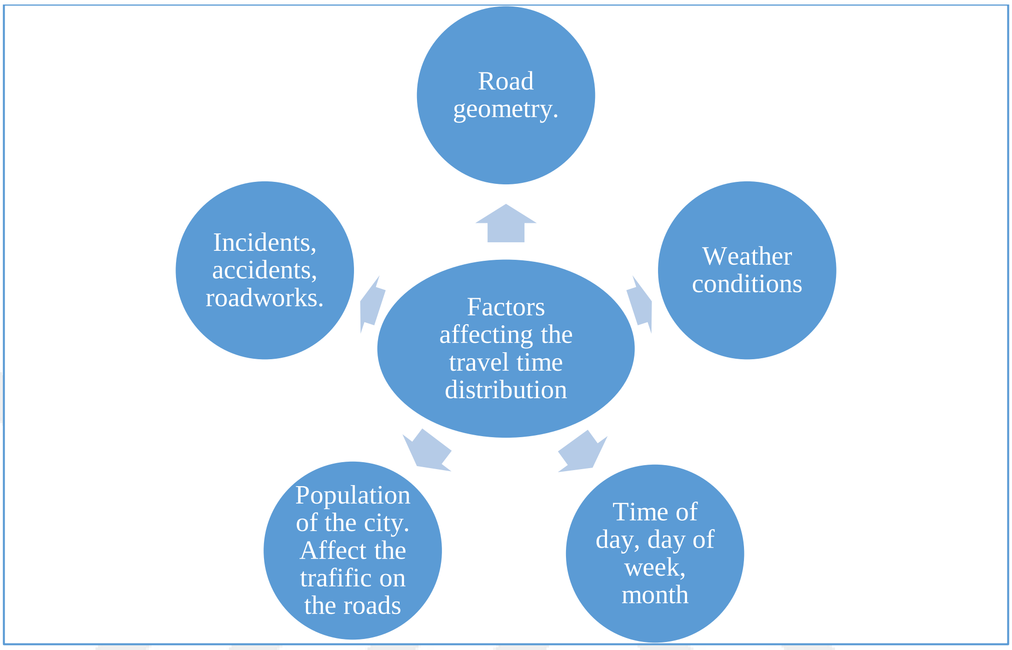

Figure 1.1 Schematic overview of factors affect the travel time distribution

As the data source, three data types are used to determine traffic conditions; the number of vehicles passing through the fixed point sensors, track based sensors and camera based sensors. Fixed point sensors are usually used in the inductive loop. They are placed in certain points on the road are used to count the number of passing vehicles from those points. Loop detectors that are established in two points on the road determine the number of vehicles between those two points and provide a measure of traffic density (Kwon et al., 2000). However, the ring of detectors can be used to measure traffic density between two points if the road is not secondary way; hence, using public urban detectors are more suitable for areas such as highways. On the other hand, loop detector installation, electrical and data connection are expensive solutions, therefore retrieving the data by means of this method costs too much.

The other technique for data source is using global positioning system (GPS) technique (Weigang et al., 2002). GPS information includes the bus speed, bus latitude, bus longitude, last transaction time and etc. These variables are retrieved by means of GPS equipment.

One of the most commonly used methods is integrating GPS receivers into vehicles. However, it may not be possible to integrate GPS receivers into all vehicles, especially in metropolitan cities. When the literature is reviewed, it is noticed that there are various approaches related to that problem. One of these approaches is using GPS units in the users' mobile phones (Park et al., 2004). In this study, instant location information of the vehicle are obtained owing to OBU (a special smart computer developed by Kentkart) installed in buses as GPS units.

Traffic density is calculated by using data gathered from cameras (Frechette & Khan, 1997). The calculation is performed by finding the number of vehicles in the image taken by cameras. This method is expensive as high-capacity data transfer is performed through cameras. Furthermore, since information is obtained from a limited number of cameras, predictions are not successful enough.

An effective method is proposed for estimating the arrival time of a bus at a given bus stop using global positioning system (GPS) data which is collected in Manisa, Turkey. When all the methods are compared, it is observed that predictions performed by using GPS information of the vehicle become more successful. As GPS data is transferred to the system once in every 30 seconds, it is possible to make dynamic predictions.

1.2 The Goal of the Study

The goal of this study is to estimate the arrival time of buses at bus stops. Benefits of knowing the estimated arrival time of buses at bus stops for operators and passengers are explained in the following:

- It gives reliable information on travel time for passengers and operators.
- It helps operators to manage the schedule of buses.
- Operators will be informed when the bus will arrive at its final destination.
- Passengers will know how long they will wait for a bus at the given bus stop.
- Passengers will be able to modify their plan if the bus will be delayed owing to traffic conditions.
- Passengers will be able to check the bus status on a web page, led panel or mobile devices. The system refreshes the estimated time continuously.
- Passengers will receive information about when the next bus will arrive at the given bus stop.
- Passengers will be able to choose other modes of transportation routes if the bus is delayed for a long time or they miss the current bus.
- It minimizes traffic congestion at pick-up and drop-off areas.
- It assists operators in managing their own schedule. For example, if the bus is delayed especially in the evening, they can change the scheduled time for these buses.

CHAPTER TWO
RELATED WORKS

2.1 Overview

Many prediction models have been implemented for traffic state estimation such as traffic density of city, bus arrival time at bus stop and total travel time through the travel. The five most commonly used methods are historical data based models, time series model, regression models, Kalman filtering model and machine learning models.

According to those methods, when the real time traffic data is used for estimation, more efficient results are obtained. The estimation methods which get attention most in instant traffic operations are Kalman filtering, neural network and a combination of time series and Kalman filtering.

2.1.1 Model Based on Historical Data

This type of prediction model derives the current and future travel time of the bus from the historical data of the previous travel at the same time period. The current traffic state is assumed the same as the previous traffic state at the same time.

Researchers indicate that when the traffic conditions is stationary in daily and weekly time period, it is reasonable to estimate of future traffic conditions from historical averages of conditions (Williams & Hoel, 2003). However Lin and Zeng assume that this model does not infer more reliable results in different regions because large city has no stationary traffic conditions and every regions of the city has different condition in daily time period (Lin & Zeng, 1999).

2.1.1.1 Using Average Travel Time

This model use the historical average travel time and other input variables and conditions to give estimation arrival time at the bus stop. Chung and Shalaby developed an expected time of arrival (ETA) statistical model using explanatory variables (Chung & Shalaby, 2007).

Two methods are used to estimate arrival time at the bus stops

- Historical data
- Current day condition. (weather condition, schedule time)

This model is developed for school buses. In contrast to transit vehicles, school buses have one route per day and have schedule time therefore they use small historical data for estimation.

The study of Chung and Shalaby; historical data, planned time (schedule time) of school buses and weather condition was used for prediction at bus stops (Chung & Shalaby, 2007).

2.1.1.2 Using Average Speed

This model uses average speed of vehicles for estimation of travel time. It estimates travel time by using data collected via GPS technology. This model is generally used for map matching techniques which can be found on a geographic information system that estimates vehicle position and travel time.

Weigang developed a model to estimate bus arrival time at bus stops using GPS information, which was implemented in the SITCUO-Information System for Urban Bus Transportation, in Brasilia (Weigang et al., 2002). The main algorithm is to estimate the travel time at the bus stops, however there is not any information about route of the buses. Therefore route is found from current location of buses before estimation at the bus stops. The position of the bus should be match with any point on route segment or straight line. Estimation of travel time parameters is the location of bus, route of bus and speed of bus. If the bus stops, the arrival time is infinite therefore it cannot be calculated. In order to overcome the problem, historical bus travel speed and current speed of the bus are used for estimation at bus stops. The actual position and results of algorithm are compared and the mean error is found to be %8. Therefore, the results indicate that these errors can be reduced when the number of segments increases in the routes.

2.1.2 Time Series Model

Time series model algorithm makes predictions for the future by making use of historical data. They assume that an external factor does not affect dynamic system in time period and current traffic condition is the same in the future.

Chien et al., indicate on the accuracy of time series models is a function of the similarity between the real-time and historical traffic patterns (Chien et al., 2002). Differences in traffic time or changes in historical and real-time travel time cause unstable estimation results, however these models have not been used for estimation of traffic volume or travel time.

2.1.3 Regression Models

Regression models examine the relation between variables. Some variables affect the estimation however some of them are not important. In contrast to historical data model, this model study on unstable traffic conditions. These models predict and explain a dependent variable with a mathematical function formed by a set of independent variables (Chien et al., 2002).

In general, regression model assist how a dependent variable affects the fluctuation when the independent variable change in time period. The regression model arises which variable is important for the estimation algorithm.

Patnaik used regression models to estimate bus arrival times. Data is retrieved from automatic passenger counter (APC) (Patnaik et al., 2004). The variables which are used for estimation are distance, number of stops, dwell times, boarding and alighting passengers and weather condition. In regression model design, some variables are static and the others constantly change. Therefore, according to the study of Patnaik, weather condition is a less important variable for estimation.

2.1.4 Kalman Filtering Model

Kalman filter has always been an inseparable part of thousands of military and civil navigation the systems developed since the first presentation in 1960. In this perspective simple, recursive digital, algorithmic filter used the instant data in order to get the general performance of the system. In order to estimate that instant value of the system variables (such as time of arrival to the specified location), the filter performs new estimations using a statistical method by weighing each new measurement in a suitable way according to the former information. The filter can also be used for detecting the current uncertainties performed for real-time quality evaluations or offline system design studies. When Rudolf Kalman presented the Kalman Filter officially for the first time in 1960, the algorithm was not welcomed well. Practical performance of the filter requires paying attention on a suitable statistical modeling and numeric sensitivity. Field of use follows:

- Navigation
- Vehicle Tracking
- Flow dynamics
- Estimation of population
- Estimations of time
- Estimation of several non-linear data.

Wall and Dailey study on estimation of vehicle arrival time (Wall & Dailey, 1999). The algorithm used historical data and real time automatic vehicle location data for the estimation. The algorithm consists of two parts such as monitoring and prediction. The Kalman filter model used vehicle movement of buses for estimation of arrival time. They did not deal with dwell time as independent variable. Their algorithm error rate is less than 12 % (Wall & Dailey, 1999). The algorithm presented to user a web application.

Shabalaby and Farhan implemented the prediction algorithm with Kalman filtering model (Shabalaby &Farhan, 2004). They used data that collected of four buses data with automatic passenger counter data and automatic vehicle location data in real time. They studied on estimation of arrival times with five weekday data. The first four weekday data was used for developed model, the last weekday data was used for testing. Their prediction algorithm worked with run time between links that contain two and eight bus stop that was check point. Dwell time did not calculate for every bus stop, they calculated only check point. They updated the prediction time of bus where the bus arrived check point. The route has twenty bus stops; according to algorithm prediction time is updated only five check point. Consequently the model demonstrate dynamic ability, update itself based on new data. Every new data reflect the prediction time.

Chien and Kuchipudi developed a travel time estimation using with real time historic data (Chien & Kuchipudi, 2003). They used Kalman filtering technique because of the state update continuously with changing condition. They comprised two different approaches; these are path base travel time values and link based travel time values. The results demonstrate that historic path base travel time is better than travel link because of especially peak hours error rate is smaller and smaller travel time variance. The advantage of historic path base estimation is prediction of travel time any given time. Disadvantage of historic path base estimation is slower than link base estimation. They are focused the estimation time accuracy of special time for example; after concerts, football match, peak hour etc.

2.1.5 Machine Learning Model

Machine learning model examines of pattern recognition and computational learning theory in artificial intelligence. Machine learning investigates the study which can learn from and make estimation on data. Machine learning model algorithms deal with complex relationships between predictors.

Hoogendoorn and Van Lint are able to operate in non- linear relationships. These models can be used for prediction of travel time, without explicitly addressing the (physical) traffic processes (Hoogendoorn & Van Lint, 2008). This model is generally used for correlation analysis, genetic algorithm and trial and error procedures. The results obtained from one point are not transferred to the next step. Machine learning model methods are Artificial Neural Network and Support Vector Regression.

Chen et al developed a method for estimation of bus arrival time using automatic passenger counter data (Chen et al., 2004). Their model consisted of Artificial Neural Network model for prediction bus arrival time and dynamic prediction model based on Kalman filter. Bus location data was used for estimation of bus travel time. ANNs model was trained with input variables that are day-of-week, time-of - day, weather and segment. In other words, ANN model was trained with four variables that were influencing the result. At the same time, it was updated by using real-time automatic passenger counter data. The results showed that dynamic algorithm model was better in performance than artificial intelligence model due to integrating the last GPS data and bus arrival information into the prediction. As a consequence, ANN model is used for non – linear relationships between travel time and independent variables.

CHAPTER THREE
DATA COLLECTION

3.1 Data Collection

The data used for this study have been collected from using bus computer device which is placed into buses in Manisa. Nine different routes have been used in this study. The length of routes is approximately 7 km, spanning 35 bus stops in each direction and the total number of buses is 202. Every bus has GPS (Global Positioning System) receiver device. When the bus arrives at bus stop or departs from bus stop, the bus location and bus stop information is sent to the system also GPS data is sent in every thirty seconds.

GPS data consists of the following attributes:

- Route of the bus
- Point locations in latitude – longitude pairs.
- Speed of the bus
- Sent GPS time
- Arrival time at bus stop and departure time from bus stop
- The last arrival at bus stop or departure from bus stop.

In this study, the historic data of Manisa between January, 2014 and July, 2014 are used as if it is real-time to perform the estimation of the developed algorithm.

3.1.1 Real-time Data Flow

Figure 3.1 shows the transition process of real time data by using bus computer equipment in the buses. The study uses a seven - day GPS information of the system. All information is stored in MYSQL database and all transactions are performed in this database.

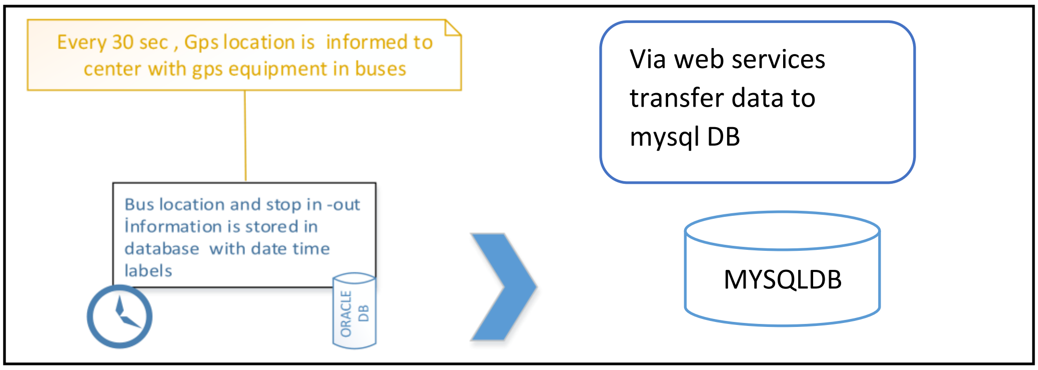

Figure 3.1 GPS information of bus

3.1.2 Static Data Flow

Routes, route-path, path bus stop data is updated every day. If any of the information changes, then the system is updated automatically. Figure 3.2 shows the flow of the static data.

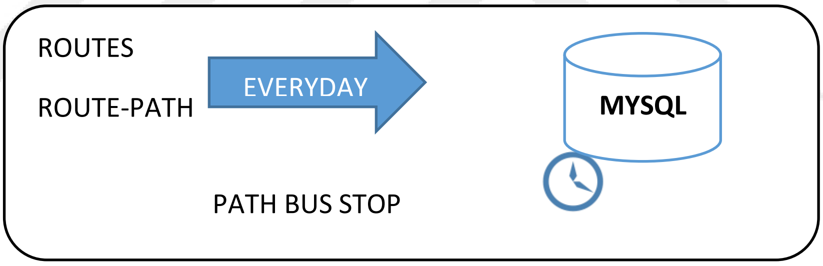

Figure 3.2 Static data of the system

3.2 Preliminary Analysis

In analysis process, transportation terminology is examined.

- **Route:** Route is a line which has a start point and end point.

- **Path**: Road segments are defined as a path. Route has more than one path. Except for forward and backward path of route, there could be any other paths.

- **Bus Stops:** Stop points. Bus stop id is unique in the system even bus stop ids are on the same route. For example, route code is 255i and the route has two directions. If both directions of the route have the same bus stop id, there becomes a trip identification conflict.

- **Bus Stops of the Path:** Stops that belong to path. Every path has bus stops with their sequence number.

- **Trips:** Traveling which starts at one point and ends at another point or the same point. It is planned in specified time periods. Trip includes attributes below:

  - **Bus Id:** It is assigned bus number in trips
  - **Trip start date time:** It is start date time of the trip.
  - **Trip end date time:** It is end date time of the trip.
  - **Trip No:** Trip unique number. Every trip has a trip number. The numbers are distinct from each other.
  - **Route Code:** Route code of the trip.
  - **Path Code:** Path code of the trip.
  - **Arrival date time at bus stop:** Arrival date time of bus at the last bus stop.
  - **Departure date time at bus stop:** Departure date time of bus from the last bus stop.
  - **Arrival bus stop id:** Id of bus stop that bus arrives at the last bus stop.
  - **Departure bus stop id:** Id of bus stop that bus departs from the last bus stop.
  - **Latitude:** The last location information of bus as latitude.
  - **Longitude:** The last location information of bus as a longitude.

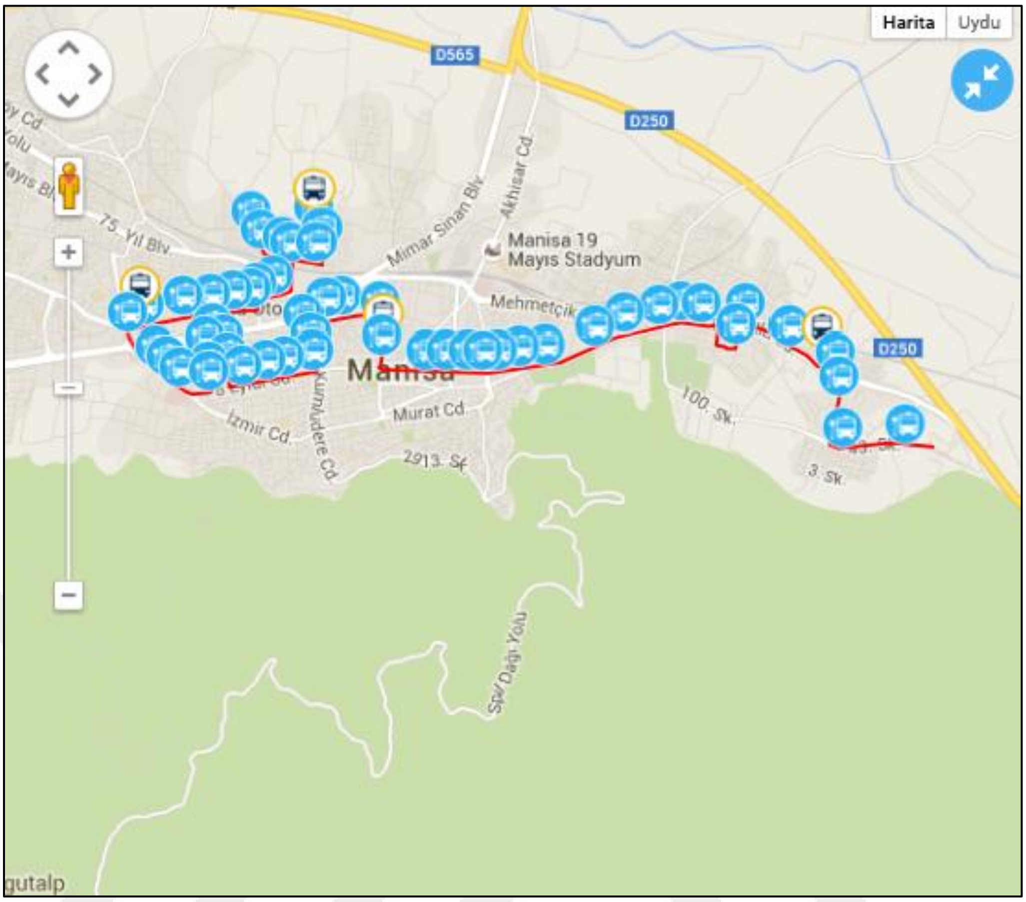

Figure 3.3 Bus stops of the path.

The Figure 3.3 is obtained by plotting the coordinates of the GPS data points which are signed in red. Blue icons are bus stops throughout the path. Bus stops of path are sequentially mapped. When the sequence number of the bus stop in the path is 1, it means the bus is at the start point of the trip. All bus stop ids are unique in the system. Opposite directions of the route have different bus stop ids.

Trips must take place in the correct order of bus stops. If the bus skips some bus stops, it means something went wrong. Possibilities are as follow:

1. Trip is assigned the other bus , therefore trips is uncompleted
2. As real-time data is sent to system with UDP (User Datagram Protocol) packets, it may be lost.

If the travel time is more than expected between a bus stop and the next bus stop, possibilities are as follow:

1. Incident occurs.
2. Traffic is busy.

3.3 Preliminary Analyze 2 Formulas

- **Travel Time:** Travel time between any two stops k & l for a journey ‘p’ is initiated in real time ‘t’ of a day ‘d’ is the is defined as the difference between departure time of bus from bus stop ‘k’ and departure time of bus from bus stop l and is defined in equation (3.1). Sequence number of ‘l’ must be greater than sequence number of k. Travel time is calculated by using departure time information of the last bus.

$$T_{l-k}^{tdp} = T_{Dl}^{tdp} - T_{Dk}^{tdp} \tag{3.1}$$

- **Bus Stop Pair:** Bus stop pair is defined as sequential bus stops. Bus stop pair sample is 10321 and 10329. Their sequence numbers are 10 and 11 of the path as 24491.

- **Sequence number:** Sequence number is dependent on path code.

- **Dwell time:** Difference between departure time from bus stop and arrival time at bus stop of the bus. Dwell time is defined in equation (3.2).

$$DT_{l}^{tdp} = T_{Dl}^{tdp} - T_{Al}^{tdp} \tag{3.2}$$

T_{l-k}^{tdp} the travel time between stop l & k for a journey p initiated at real time of a day d.

T_{Al}^{tdp} arrival time of bus at bus stop l.

T_{Dl}^{tdp} departure time of bus ar bus stop l

T_{Ak}^{tdp} arrival time of bus at bus stop k.

DT_{l}^{tdp} dwell time of bus at bus stop l

In order to calculate travel time between sequential bus stops, the difference of departure times in sequential bus stops are calculated.

Table 3.1 Departure time of buses at bus stop sequence 22

Bus	**Bus Stop Sequence**	**Departure Time**
255	22	08:56:05
297	22	08:55:00
266	22	08:57:03

Table 3.2 Departure time of buses at bus stop sequence 23

Bus	**Bus Stop Sequence**	**Departure Time**
255	23	08:59:25
297	23	08:58:56
266	23	08:59:03

Table 3.1 shows departure time of buses at bus stop sequence 22. Table 3.2 shows departure time at bus stop sequence 23. Travel time is calculated according to departure time of Table 3.1 and Table 3.2. Travel time of three buses is following:

- Travel time 120 sec for bus as 266.
- Travel time 200 sec for bus as 255.
- Travel time 234 sec for bus as 297.

CHAPTER FOUR
SYSTEM DESIGN

4.1 System Architecture

Figure 4.1 shows the system architecture of the estimation algorithm. Data are collected by OBU in the buses. Data are stored in ORACLE database of Kentkart system. Data consist of global position system data of seven days that are transferred to MySQL database with web services. An estimation algorithm calculates estimation time of the bus at the given bus stop over by MYSQL database. As location updates arrive from the buses, they will go through bus stop. The prediction algorithm will periodically update its state from this database. The led panel or mobile application as a tracking application, which will act as the user interface, will listen the requests for a specific bus stop, call the prediction algorithm and return the results as a response.

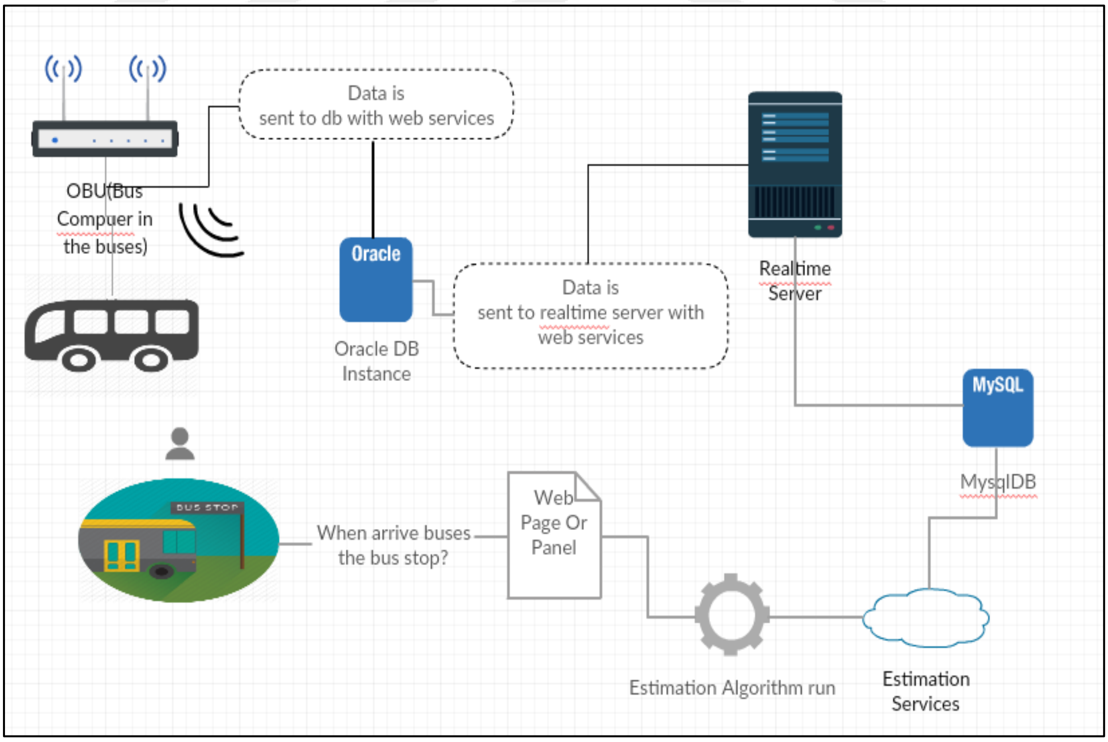

Figure 4.1 Architecture of system

4.2 Database Design

MySQL database is used for configuration control, storing data and consists of seven tables (see Figure 4.2) describing all the parameters of the system:

- **Routes:** It is master table, route_id is the primary key.
- **Route_path:** Route_id and path_id is foreign key.
- **Path_stop:** Path_code and bus_stop_id is foreign key.
- **Stops:** It is master table stop_id is primary key.
- **Paths:** It is master table and path_id is primary key.
- **Bus_gps_data:** It is updated in ten seconds, bus id is primary key
- **Bus_stop_pair:** Pair1 and pair2 is primary key.

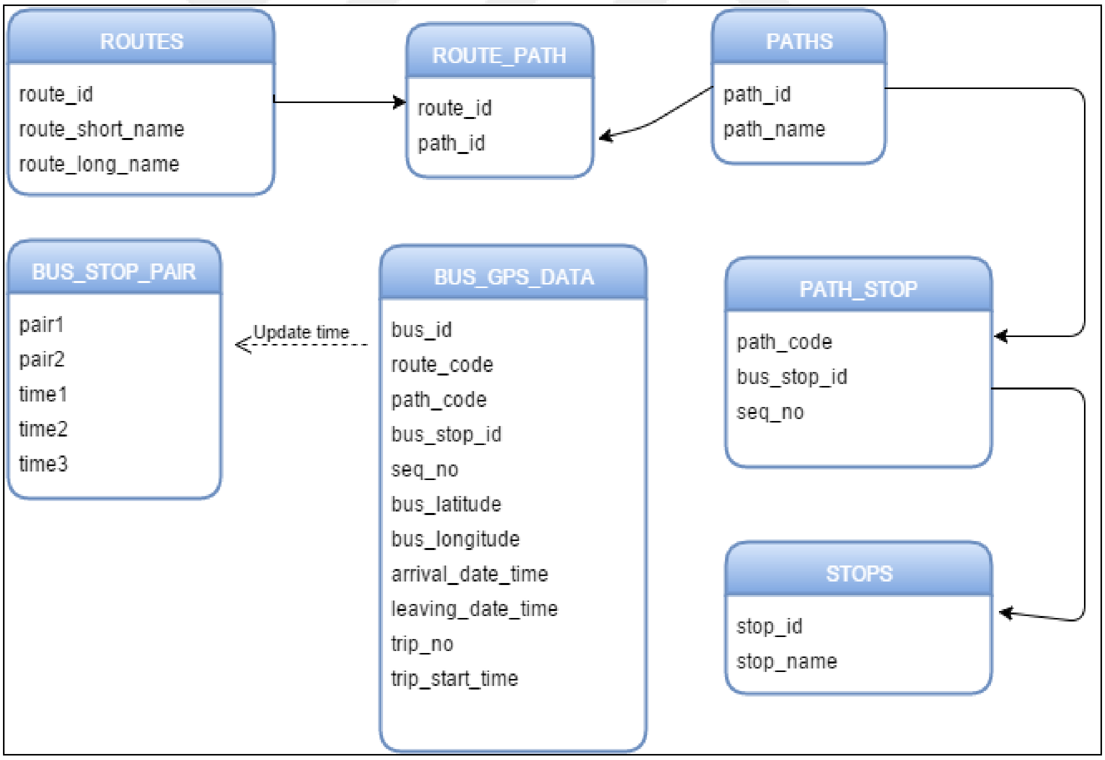

Figure 4.2 Database design

4.2.1 Routes

All route information is stored in "Routes" table. A route is a group of trips that are displayed to riders as a single service. Table 4.1 shows description of Route table's columns.

Table 4.1 Description of routes table

FIELD NAME	FIELD DESCRIPTION
ROUTE_ID	The route_id field contains an ID that uniquely identifies a route. The route_id is dataset unique.
ROUTE_SHORT_NAME	The route_short_name contains the short name of a route. This will often be a short, abstract identifier like "255", "4" that riders use to identify a route, but which doesn't give any indication of what places the route serves..
ROUTE_LONG_NAME	The route_long_name contains the full name of a route.

4.2.2 Path

Path is defined as a direction of route. Rules for drawing lines on a map represent a transit organization's routes. Table 4.2 shows description of Path table's columns.

Table 4.2 Description of path table

FIELD NAME	FIELD DESCRIPTON
PATH_ID	Direction of route.
PATH_NAME	Path name of the route. The path name contains the full name of a path.

4.2.3 Route Path

Route path table provides relation between routes and path. Table 4.3 shows description of Path table's columns

Table 4.3 Relation between route and path

FIELD NAME	FIELD DESCRIPTON
PATH_ID	This value is referenced from path table.
ROUTE_ID	This value is referenced from route table.

4.2.4 Path Bus Stop

Path bus stop table represent bus stops which belongs to a path. Table 4.4 shows description of Path Bus Stop table's columns.

Table 4.4 Relation between Path and bus stop

FIELD NAME	FIELD DESCRIPTION
PATH_CODE	Path id of the route
BUS_STOP_ID	The bus_stop_id field contains an ID that uniquely identifies a stop. The stop is referenced from the stops table.
SEQ_NO	The seq_no field identifies the order of the stops for a particular path. The values for seq_no must be non-negative integers, and they must increase along the path.

4.2.5 Bus GPS Data

This table contains global positioning data of each bus and is updated every ten seconds. Table 4.5 shows description of Bus Gps data table's columns

Table 4.5 GPS information of the bus

FIELD NAME	**FIELD DESCRIPTION**
BUS_ID	Bus id of moving buses
ROUTE_ID	Route code of the bus
PATH_ID	Path code of the bus
BUS_STOP_ID	Arrival or leaving bus stop id of the bus
SEQ_NO	Arrival or leaving bus stop id sequence of the bus
BUS LATITUDE	Latitude of the bus
BUS LONGITUDE	Longitude of the bus
ARRIVAL_DATE_TIME	Arrival date time at bus stop of the bus
LEAVING_DATE_TIME	Leaving date time at bus stop of the bus
TRAVEL_TYPE	Type of bus stop arrival or departure
TRIP_ID	The trip_id field contains an ID that identifies a trip. The trip_id is dataset unique.
TRIP_START_TIME	Start date time of the trip.

CHAPTER FIVE
ALGORITHMS AND APPLICATION

Kalman filter and average travel time algorithms are used to estimate travel times from current bus stop to the target bus stop.

5.1 Kalman Filter

Kalman filter is a digital filter having multiple inputs and outputs. This filter optimally estimates the states of a system that has real-time noisy outputs (Figure 5.1). States of the system (values of each time) are variables that need to identify the entire system behavior as a function of time (such as position, speed, time, voltage level and etc.). Indeed, the system having multiple noisy outputs is supposed to have multi-dimensional noisy signal. It is not clear which of those show the state of the system. Kalman filter filters the signal that is out of noisy measurements in order to estimate. Kalman filter is shown as a more general method comparing to other acceptable methods.

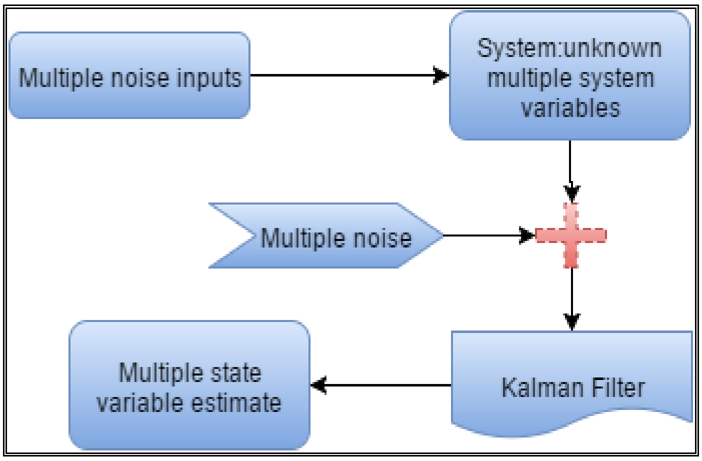

Figure 5.1 Process of Kalman Filter

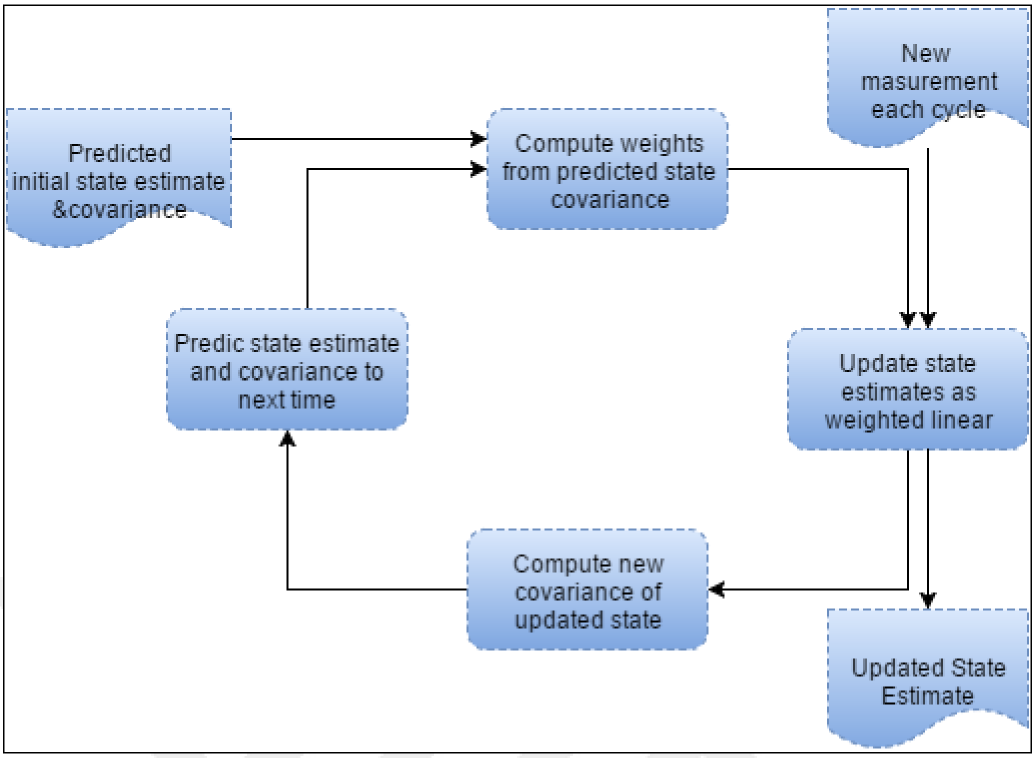

Figure 5.2. Kalman filter algorithm

Kalman Filter is a recursive linear filter. In each cycle, state estimation is updated by combining the new measurements with the predicted state estimations that are taken from former measurements.

As it is seen in Figure 5.2, estimation starts from the first value and it is combined with the covariant variable taken from former information. While evaluating the filter weighs, this estimation with the first measurement vector is used for generating the well-updated new estimation. If the noise covariant of measurement is less than the estimation, then the measurement weigh is raised up; or else, reduced down.

5.1.1 Development of Prediction with Kalman Filter

The estimation of how long it takes for a bus to arrive at the bus stop is performed by using Kalman filter. In order to understand the estimation model, Figure 5.3 shows the example transit route. The route is divided into segments. Each segment has two bus stops. When the bus leaves from the bus stop, the system stores the departure time as the actual departure time at bus stop. In this example, Kalman filter prediction algorithm for running travel time will estimate the next bus stop running travel time.

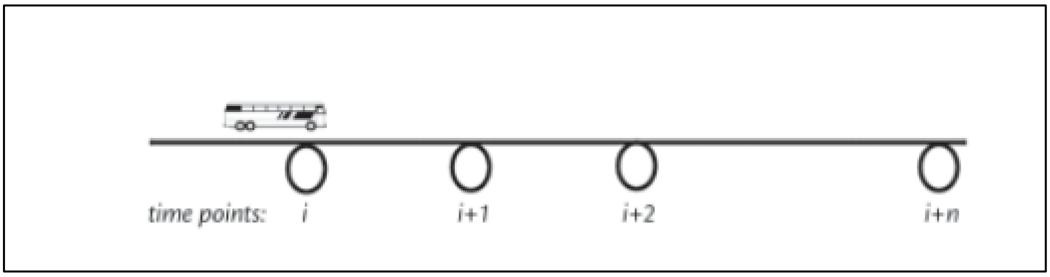

Figure 5.3 Example of the transit route

Assume that the bus is at bus stop I as follows in equation 5.1

$$At\ (i+1) = Dt\ (i) + Rt\ (i,i+1) \tag{5.1}$$

- At (i+1) is the predicted arrival time of bus n at stop i+1
- Rt (i,i+1) is the predicted running time between i and i+1 from Kalman Filter prediction algorithm
- Dt (i) is the actual departure time of bus n from stop i

Kalman filter algorithm is implemented for specified segments. The length of those segments differs between almost 100 meters and 500 meters. Kalman filter is performed for each segment. Those segments consist of different linear bus stops. Each segment is formed by bus stop couple. Linear bus stops consist of bus stops that have consecutive bus stop sequence number. There are a lot of buses passing through the same linear bus stops although they have different routes regions and numbers which increases the density of the passing vehicles there.

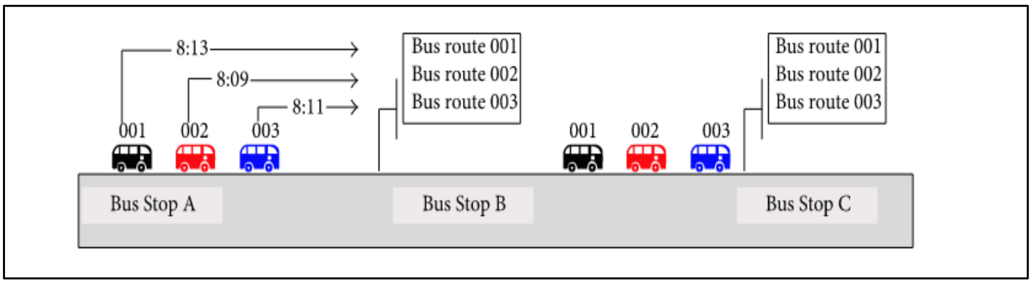

Figure 5.4 Example route and bus stops.

In order to understand for dividing into segments in routes, Figure 5.4 shows example of multiple bus routes sharing the same road segments. For example, a passenger expects to travel from B bus stop to C bus stop. The three buses will arrive at bus stop C. Therefore, the passenger chooses any of three buses. The three buses have different routes; however some target points are the same. Therefore many buses pass between bus stop A and bus stop B pass. Running travel time between B and C is independent from routes.

The rate of reliability of Kalman filter increases according to density situation of instant data. Therefore, the study uses the segments which include consecutive bus stops. Every instant data generates a new estimation and are updated with the former estimated one. Performed studies include one-day data set. Every coming bus data are used for estimation. For each arrival time that busses send, estimation is performed for the next bus stop of the one where the bus is. Each arrival time is used for performing estimation for the next bus stop. By comparing the each coming new data with their former ones, new arrival time estimation is performed owing to Kalman filter. The steps of the Kalman algorithm are as follows:

1. The first step is to calculate the Kalman gain. Kalman gain is a multiplier resulted from the previous estimation in order to make better estimations. The gain is regulated for better performance, because with using a high gain, filter follows the observations closer. With a lower gain, filter follows the model estimations closer. The method converges to perform closer estimations than the ones that are performed based on real unknown values, one-time measurements or just model estimations.

$$K_{G}=\frac{E_{EST}}{E_{EST}+E_{MEA}} \tag{5.2}$$

Where is E_{EST}error variance in estimation, where E_{MEA} is error variance is measurement. Kalman gain formula is shown in equation (5.2).

Error variance is the difference between real measurement travel time and average of all measured travel time. For every measured value, a new error variance is calculated. While calculating this deviation value, average passing time between two bus stops is used. Average passing time is subtracted from the instant passing time between two bus stops which results the error variance as it is seen in equation (5.3).

$$E_{MEA} = \text{MEA-} \sum_{1}^{n} \frac{MEA}{n} \tag{5.3}$$

2. The second step leads to the estimation of the arrival time for the next bus stop. Estimation of arrival time is calculated according to equation (5.4). In this process, previously estimated value, Kalman gain and real arrival time are used.

$$EST_t = EST_{t-1} + KG(MEA - EST_{t-1}) \tag{5.4}$$

where EST_t is the estimation travel time at the target point, EST_{t-1} is the previous travel time at the target point and MEA is the measurement time in the subsection.

3. In the last step, the new estimated error value which will be used in new estimations is found and used in equation (5.5).

$$E_{EST_t} = (1 - KG)E_{EST_{t-1}} \tag{5.5}$$

Table 5.1 Example of kalman filter algorithm

Input Values	Kalman filter example		
	k	0	1
	MEA	82	79
	$\sum_1^n \frac{MEA}{n}$	73	73
	$\mathbf{EST_t}$	70	74.8
	$\mathbf{E_{MEA}}$	9	6
Time Updating	$\mathbf{EST_{t-1}}$	70	74.8
	$\mathbf{E_{EST_{t-1}}}$	6	3.6
Measurement Updating	**KG**	$K_{G=}\frac{E_{EST}}{E_{EST+E_{MEA}}}$ =(6/(6+9))=0.4	=3.6/(3.6+6) =0.06
	$\mathbf{EST_t}$	$EST_t = EST_{t-1} + KG(MEA - EST_{t-1})$ $EST_t = 70 + (0.4) * (82 - 70)$ =74.8	75
	$\mathbf{E_{EST_t}}$	$E_{EST_t} = (1 - KG)E_{EST_{t-1}}$=(1-0.4).6 E_{EST_t}=3.6	3.38

Table 5.1 explains Kalman Filter with examples. MEA describes the difference of passing times of buses between linear bus-stops. Each passing time of buses is updated in the field of MEA. EST t indicates the estimated passing time between bus stops. $\sum_1^n \frac{MEA}{n}$ indicate that the average of the travel time between linear bus stops and calculated in one time.

5.2 Average Travel Time Algorithm

The other method is average travel time for the estimation of how long it takes for a bus to arrive at the bus stop.

As it is explained in Kalman filter, departure time of bus at bus stop is used as the realized time at the bus stop. If the new data comes to the system, the algorithm calculates the travel time at sequential bus stops of bus. The last three travel time information is stored for each sequential bus stop and is updated when the new data come. Estimated travel time is found with average of the travel time data of three buses at sequential bus stops.

5.2.1 Estimation Travel Time Service with Average Travel Time

Estimation of travel time service is calculated in MySQL database. All estimation calculations have been performed when the estimation service is called with bus stop id parameter. For example when the user uses the bus stop id as a parameter in the interface in order to get information of a bus, algorithm calculates estimation travel time for the bus stop where the user is.

Assume that the bus is at bus stop i as follows equation 5.6;

$$\text{AvgT}\ (i+1) = \text{DT}\ (i) + \text{RT}\ (i,i+1) \tag{5.6}$$

- AvgTn (i+1) is the predicted arrival time of bus n at stop i+1
- RTn (i,i+1) is the predicted running time between i and i+1 from average travel time prediction algorithm
- DTn (i) is the actual departure time of bus n from stop i

According to the 5.7 equation, if ‘i’ is the place where the bus is and ‘i+2’ is the target point;

$$\text{RTn}(i,i+1)=T(i,i+1)/3+T(i+1,i+2)/3 \tag{5.7}$$

- **T(i,i+1)** is total travel time at bus stop i and bus stop i+1 of the last three buses.
- **T(i+1,i+2)** is total travel time at bus stop i+1 and bus stop i+2 of the last three buses.

5.3 Dynamic Travel Time Calculation for Bus Stop pair

Estimation of arrival time at bus stops is based on realized arrival time and departure time information which gather from buses. Travel time of the last three buses at sequential bus stops stored in database.

All sequential bus stops are defined in the system. Every bus has arrival time at the bus stop and departure time from the bus stop with date time information, therefore the travel time can be calculated in sequential bus stops with specified time range.

Table 5.2 GPS data of Bus

BUSID	PATH_ID	STOPID	SEQ_NO	ARRIVAL_T	LEAVING_T
12345	1	100	5	10:05:10	10:05:46
12345	1	101	6	10:08:56	10:09:22
23456	1	100	5	10:08:55	10:09:25
23456	1	101	6	10:11:38	10:11:58
33333	2	100	6	10:12:11	10:12:45
33333	2	101	7	10:14:55	10:14:10
44444	3	100	12	10:14:58	10:15:01
44444	3	101	13	?	?

Table 5.2 shows arrival and departure information with bus stop data of three buses with date time values. The bus whose id is 12345 arrives at the bus stop at 10:05:10 and arrives at the next bus stop at 10:08:56. Travel time of 12345 between these two sequential bus stops is 226 sec. For 23456, travel time is 163 sec for the same sequential bus stops. For 33333, travel time is 175 sec for the same sequential bus stops. If the three buses are the last three buses for the bus stops whose ids are 100 and 101, the average of the travel time for 100 and 101 is 188 sec. The three buses' historical information affects travel time of the bus whose id is 44444, therefore the last three buses affect current arrival time estimation of a bus at the next bus stop.

Travel time calculations for sequential bus stops are updated when the new bus information is obtained.

5.4 Estimation Travel Time Service Flow

In Figure 5.5, a flowchart of finding the estimated time for required bus stop id is shown.

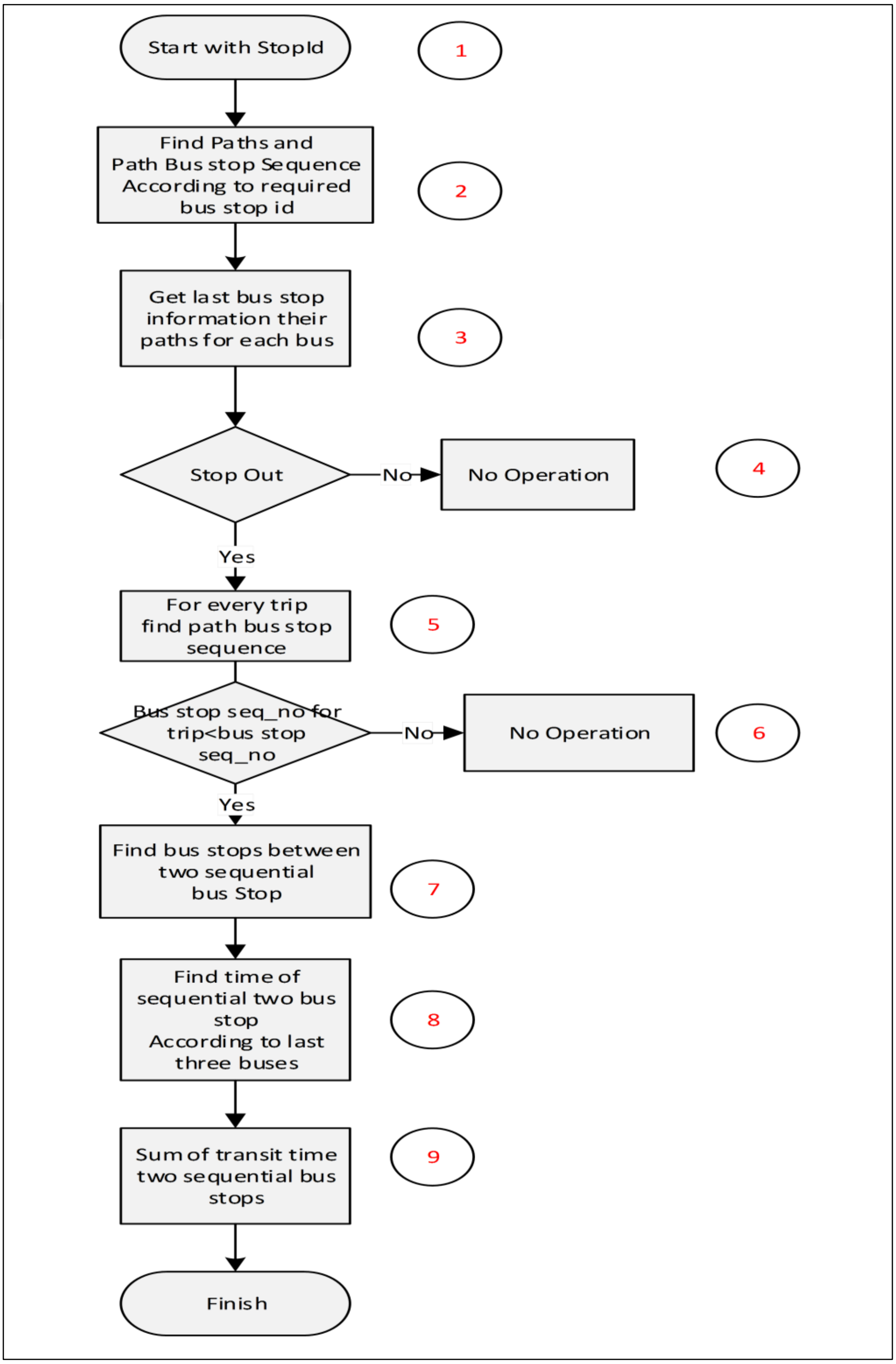

Figure 5.5 Flowchart of estimation time service

In the following information is used for finding estimation:

1. Routes Information
2. Path Information
3. Path Stop Information
4. Arrival time at the last bus stop of the bus

- **Step 1:** Service is started to run with bus stop id parameter.

- **Step 2:** Find paths and path bus stop sequence according to required bus stop id as follows in Figure 5.6.

Figure 5.6 Path Stop table

P1:Bus stop id

R1:Path Id

R2:Bus Stop Sequence

- **Step 3:** Get the last bus stop information in their path for each bus. Real time and historical data are stored in "BUS_GPS_DATA" table. Trip's start date time and bus id are used for finding the last trip.

The last GPS data of buses have to be as follows:

- Trip is not closed.
- Max trip start time
- Bus ID

- **Step 4:** Departure date time is when the bus depart from bus stop is used for all calculations. All estimations are based on departure time at the bus stop of the buses. Therefore, dwell time is not used for estimations.

- **Step 5:** Bus stop sequence of path is calculated for each bus in every last trip. When bus stop sequence is calculated, “BUS_GPS_DATA” and PATH_STOP tables are used. Path bus stop sequence is found for each bus and only calculated for the last trips and unclosed trips.

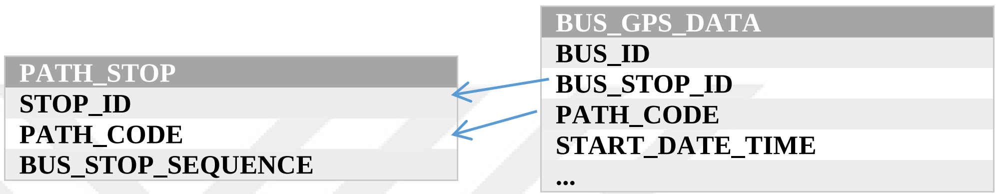

Figure 5.7 Relation between path_stop table and bus_gps_data table

Figure 5.7 shows the relationship between path bus stop table and bus GPS data table.

- **Step 6:**
  - Path bus stop sequence is calculated for required bus stop in second step.
  - Path bus stop sequence is calculated according to bus stop information of trip for each bus in the fifth step.
  - Two path bus stop sequence numbers are compared in this step. Bus stop sequence number of trip has to be smaller than required bus stop sequence number because bus will come to required bus stop and is behind the required the bus stop.

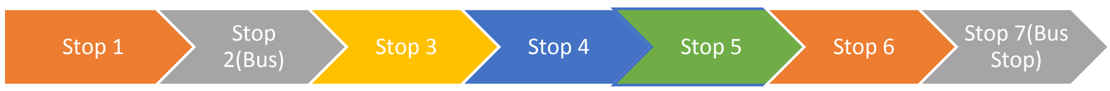

Figure 5.8 Example step 6

Figure 5.8 shows an example of step 6. Bus departs from bus stop sequence two and travels towards target bus stop 7. There is five bus stops left.

- **Step 7:** Find bus stops between two sequential bus stops. Path bus stop table and bus_gps_data tables are used. Sequence numbers are calculated in step 6, therefore between two bus stops sequence numbers are calculated. Figure 5.9 shows the remained bus stop sequence number. For example, for path 1, sequence number 3,4,5,6 are between two sequential bus stops shown as follows:

    - Stop 2-3 (correspond to bus stop id 123 124)
    - Stop 3-4 (correspond to bus stop id 124 125)
    - Stop 4-5 (correspond to bus stop id 125 126)
    - Stop 5-6 (correspond to bus stop id 126 127)
    - Stop 6-7 (correspond to bus stop id 127 128)

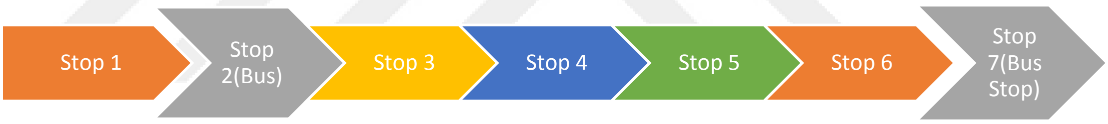

Figure 5.9 Example step 7

- **Step 8:** Bus stop ids are calculated for two sequential bus stop in step 7.

    - Stop 2-3 (correspond to bus stop id 123 124)
    - Stop 3-4 (correspond to bus stop id 124 125)
    - Stop 4-5 (correspond to bus stop id 125 126)
    - Stop 5-6 (correspond to bus stop id 126 127)
    - Stop 6-7 (correspond to bus stop id 127 128)

Transition time is calculated for bus stop pairs. This data is retrieved from bus_gps_data table. Table 5.3 shows the transition time for sequential bus stop pair in the last three buses. Time diff means the difference in time between bus stop 123 and 124. This process is performed for every calculated bus stop pair.

Table 5.3 Transit time of the sequential bus stops of the buses

Bus	TripTime	DTime123	DTime124	TimeDiff
12345	17:00:12	17:00:12	17:04:12	252
12356	17:10:12	17:18:12	17:21:12	180
12347	17:20:12	17:30:12	17:33:52	220

- **Step 9:** In step 8, time difference has been calculated for every bus stop pairs. In step 9, arithmetic average of time difference is calculated for last three buses.

  According to the following example:
  (time1+time2+time3) /3→(252 + 180 + 220)/ 3=217.3

- **Step 10**: The sum of time difference is calculated for each remained bus stop pair.

For stop 2-3 (correspond to bus stop id 123 124), bus arrives from 123 to 124 approximately in 217 seconds. This calculation process is repeated for every bus stop pair.

- Stop 3-4 (correspond to bus stop id 124 125) (approximately 205 second)
- Stop 4-5 (correspond to bus stop id 125 126) (approximately 210 second)
- Stop 5-6 (correspond to bus stop id 126 127) (approximately 455 second)
- Stop 6-7 (correspond to bus stop id 127 128) (approximately 305 second)

Summation of the bus stop pair → 217 + 205 + 210 + 455 + 305 = 1392 second.

Table 5.4 shows that estimation arrival time at bus stop of the bus.

Table 5.4 Estimation time of the bus

BUS	PATH	ARRIVAL_DATE_TIME	TIME_DIFFERENCE	ESTIMATION_TIME
35555	207010	20140701170012	1392	20140701172324

CHAPTER SIX

EXPERIMENTAL RESULTS

Prediction models have been developed with Kalman filter and arithmetic average time algorithm. It is important to evaluate them in terms of estimation accuracy. The study focuses on Kalman filter and arithmetic time average algorithm. Kalman filter prediction performance is compared with the arithmetic average time. Estimation samples are taken from specified time range and specified regions. Seventeen bus stop pairs are chosen for estimation. The number of vehicles in these bus stops is more than the one in other bus stop.

6.1 Tests of Average Travel Time Algorithm

Figure 6.1 shows the estimation time and realization time at the given bus stops. Estimation algorithm is performed for the random ten bus stops.

	BusId	RouteCode	StartDateTime	PDate	EstimationTime	RealizedTime	TimeDiff	StopDiff
2	16413	20151	20130812110748	12.08.2013 00:00	20130812112137	20130812112130	7	1
3	16413	20151	20130812110748	12.08.2013 00:00	20130812112144	20130812112130	14	2
4	16413	20151	20130812110748	12.08.2013 00:00	20130812112042	20130812112130	-47	2
5	16413	20151	20130812110748	12.08.2013 00:00	20130812112106	20130812112130	-24	3
6	26571	20201	20130812105015	12.08.2013 00:00	20130812111909	20130812111800	9	2
7	29329	29801	20130812112141	12.08.2013 00:00	20130812114551	20130812114611	-19	2
8	29329	29801	20130812112141	12.08.2013 00:00	20130812114550	20130812114611	-20	2
9	29329	29801	20130812112141	12.08.2013 00:00	20130812114526	20130812114611	-44	2
10	29329	29801	20130812112141	12.08.2013 00:00	20130812114525	20130812114611	-45	3
11	29329	29801	20130812112141	12.08.2013 00:00	20130812114530	20130812114611	-40	4
12	29329	29801	20130812112141	12.08.2013 00:00	20130812114433	20130812114611	-37	4
13	29329	29801	20130812112141	12.08.2013 00:00	20130812114530	20130812114611	-40	5
14	29329	29801	20130812112141	12.08.2013 00:00	20130812114531	20130812114611	-39	6
15	29329	29801	20130812112141	12.08.2013 00:00	20130812114456	20130812114611	-14	7
16	29329	29801	20130812112141	12.08.2013 00:00	20130812114358	20130812114611	-12	7
17	29329	29801	20130812112141	12.08.2013 00:00	20130812114323	20130812114611	-47	8
18	29329	29801	20130812112141	12.08.2013 00:00	20130812114402	20130812114611	-9	9
19	28249	29861	20130812110535	12.08.2013 00:00	20130812113045	20130812113023	22	1
20	28249	29861	20130812110535	12.08.2013 00:00	20130812113024	20130812113023	1	2
21	28249	29861	20130812110535	12.08.2013 00:00	20130812112951	20130812113023	-31	2
22	28249	29861	20130812110535	12.08.2013 00:00	20130812112955	20130812113023	-27	3
23	28249	29861	20130812110535	12.08.2013 00:00	20130812113012	20130812113023	-11	3
24	28249	29861	20130812110535	12.08.2013 00:00	20130812113005	20130812113023	-18	4
25	28249	29861	20130812110535	12.08.2013 00:00	20130812112953	20130812113023	-29	4

Figure 6.1 Experimental test results

Figure 6.1 shows the estimation time and the realization time at the given bus stop.

- Estimation time column describes the estimation time of the bus at the given bus stop.
- Realized time column describes the realized time of the bus at the given bus stop.
- Time diff column describes the difference between estimation and realized time of the bus at the bus stop.
- StopDiff column describes the number of remaining bus stop of the bus to the given bus stop.

The first record of excel in Figure 6.1 such as 16413 (bus id) arrives at the bus stop 7 seconds earlier then the estimation time when the number of remaining bus stop is one. The error rate which is defined as difference realization time and estimation time is 7 second for this example. The error rate is 14 sec for the second record of excel. In this way, the error rate calculation is calculated among 8000 dataset which consist different number of remaining bus stops which is start from one to ten. Therefore accuracy rate is revealed for estimation at bus stops. The Figure 6.2 is composed from the test results. This figure shows difference time between realization time and estimation time with number of remaining bus stop. According to the results show that the fewer number of remaining bus stops to the given bus stop, the more efficient forecasting accuracy is performed.

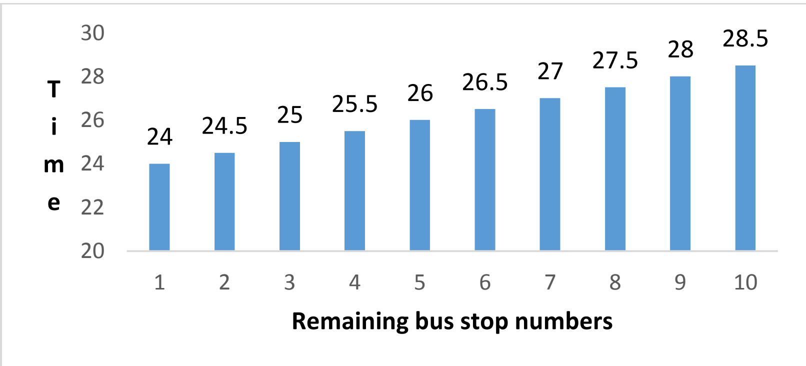

Figure 6.2 Difference between realize and estimation time

Figure 6.3 shows realized time and estimation time of the bus according to remaining to bus stop number. In the figure, x axis shows the remaining bus stop number of the buses and y axis for blue icons show estimation time, for red icons show realized time. According to the figure, the bus arrives at the bus stop (number 1) approximately in 65 seconds; however the realized time of the bus is 90 second. The error rate is 25 second for this example. When all the results are examined among 8000 data, the error rate is found between 25 second and 30 second. This results show the accuracy rate for estimation is successful.

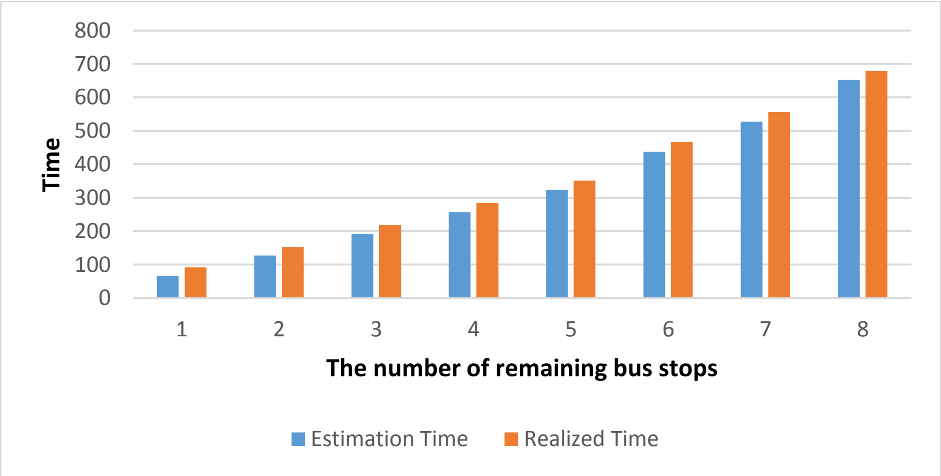

Figure 6.3 Estimation time and real time compare graph

6.2 Compare of Average Travel Time Algorithm and Kalman Filter

Kalman filter and average travel time algorithm are compared with test cases. Time periods and bus stop pairs are specified for performance tests.

Table 6.1 Compare table with Kalman Filter, Arithmetic average time and Real time

StopPair1	StopPair2	Kalman	Arithmetic Avg	Real Time
80037	80337	87	80	89
80041	80040	63	66	56
80041	80039	82	92	98
80042	80041	20	18	32
80045	80041	107	85	120
80216	80217	680	752	782
80217	80218	61	58	75
80336	80337	141	125	94
80337	80867	173	130	183
80337	80345	175	204	181

Table 6.1 shows estimation travel time and realized travel time for specified bus stop pairs. Estimation travel time is calculated with Kalman filter algorithm and average travel time algorithm.

The mean absolute percentage error (MAPE) is used for comparing their performance. Performance evaluation is achieved for specific time range. Time range is determined according to rush hours of a day. Tests have been performed between 07:00 - 08:00 am, 01.00 – 02:00 pm and 05:00 – 06:00 pm. The MAPE is shown in equation (6.1) as follows.

$$\text{MAPE}=\frac{1}{n}\sum_{i}^{n}\left(\frac{|ye-yr|}{ye}\right)*100\% \qquad (6.1)$$

- Ye is the estimation travel time from current bus stop to target bus stop.
- Yr is the realized travel time from current bus stop to target bus stop.
- n is the number of test sets.

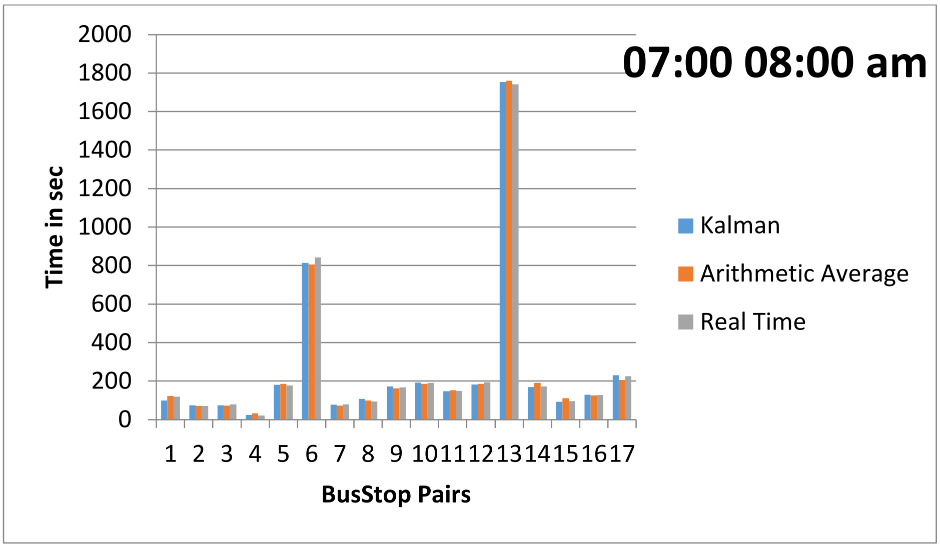

Figure 6.4 Compare table 07:00 - 08:00 pm

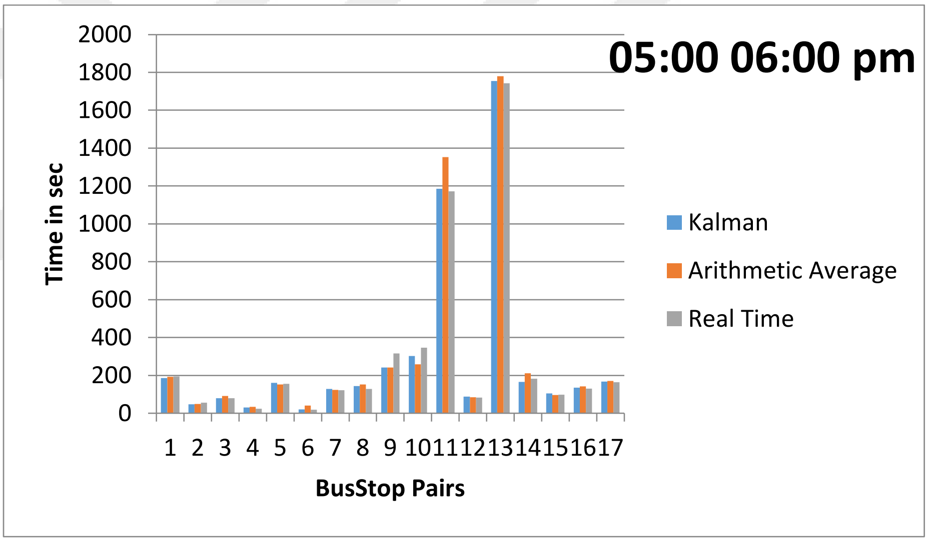

Figure 6.5 Compare table on 05:00 06:00 pm

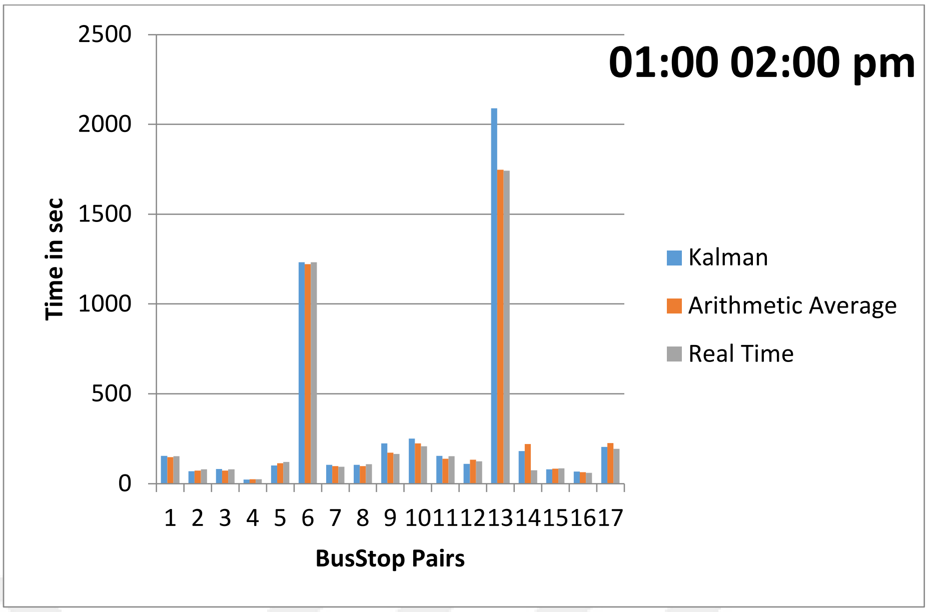

Figure 6.6 Compare table 01:00 - 02:00 pm

Tests are performed during peak hours and peak off hours. Peak hours are determined in our test cases as 07:00 – 08:00 am and 05:00 – 06:00 pm. The mean absolute percentage error (MAPE) is used for both models which are compared with MAPE in their time sections.

First time section 07:00 – 08:00 am is the morning rush hour in Turkey. This has been illustrated in Figure 6.4. MAPE value is 4.83% for Kalman filter, 6.83% for arithmetic average time algorithm. The results reveal that Kalman filter has more reliable results than arithmetic average algorithm.

The second section is 05:00 – 06:00 pm as evening rush hour. This has been illustrated in Figure 6.5. MAPE value is 8.19% for Kalman filter, 14.11% for arithmetic average time algorithm. The result reveals that Kalman filter has more reliable results than arithmetic average algorithm in rush hours.

The other test sample is 01:00 – 02:00 pm time section. This has been illustrated in Figure 6.6. MAPE value is 12.89% for Kalman filter, 9.79% for arithmetic average time algorithm. Arithmetic average is more reliable than Kalman filter in this time section.

CHAPTER SEVEN
CONCLUSION AND FUTURE WORKS

7.1 Conclusion

The goal of the study is to estimate the arrival time of buses at bus stops. There are some models that investigate the estimation time at bus stops. There is not any travel time prediction method classified as ‘commonly used’ by researchers, yet they offer some satisfactory solutions. The most successful method of all is the one which uses real time data which is the retrieval information of the buses such as speed, arrival time to bus stop, departure times from bus stop and etc.

The GPS data helps us estimate the arrival time of the bus to next bus stop. GPS - based methods are faster and cheaper. Owing to the GPS data gathered from buses, we can obtain a large number of observations containing the positions, speeds, arrival time, departure time to bus stop of various buses. Historical data is used as a real time data in estimation process Historical data consist of speed of bus, location, arrival or departure time to bus stop of bus.

Kalman filter and average travel time algorithms are performed in the study. In both algorithms, the passing times of vehicles between bus stops are supposed to be real passing times and estimations are performed based on those times. By Kalman Filter, all coming input / output information of a bus stop creates a new estimation value and Kalman gain, then old values are updated by those new ones. In average time algorithm, the estimation time evaluation is performed based on the passing time of the last three vehicles between bus stops.

Tests for Kalman and average time algorithm are performed separately. When the average time is evaluated as single, the average deviation is approximately 24 seconds. The amount of deviation increases by the number of bus stops to be estimated. According to the comparison of Kalman Filter and average time algorithm, it is observed that Kalman Filter is more successful in the days when the traffic is intense and the hours when the number of trips is high. The error ratio of Kalman Filter is 4.83% and 12.89%.

7.2 Future Works

This study investigates the dynamic travel time estimation models for buses. Only GPS data of the buses is collected for estimation at bus stops. Future works will be to collect much more data, and more factors such as the weather condition, national holidays and the travel times of other type vehicles should be considered in the models.

REFERENCES

Sun, D., Luo, H., Fu, L., Liu, W., Liao, X. & Zhao, M.(2007). *Predicting bus arrival time on the basis of global positioning system data.* Transportation Research Record. Washington: National Academy.

Hoogendoorn, S. & Van Lints, H. (2008). Online learning solutions for freeway travel time prediction, 2008, 3. *Intelligent Transportation Systems, IEEE Transactions.*

Chien, S.I.J., Ding,Y. & Wei, C. (2002). Dynamic bus arrival time prediction with artificial neural networks. *Journal of Transportation Engineering, 128,* 5, 429-438.

Williams, B. & Hoel, L. (2003). Modeling and forecasting vehicle *traffic* flow as a seasonal arima process: Theoretical basis and empirical results. *Journal of Transportation Engineering, 129*(6):664–672.

Lin, W.H. & Zeng, J. (1999*).* Experimental study of real-time bus arrival time prediction with GPS data*. Transportation Research Record*, 1666, 101-109.

Li Weigang, Koendjbiharie W., de M Juca R.C., Yamashita Y. & Maciver A. (2002) Algorithms for estimating bus arrival times using GPS data, Intelligent Transportation Systems. *The IEEE 5th International Conference* (868 – 873).

Chung, E.-H., & Shalaby, A. (2007). Expected time of arrival model for school bus transit using real-time global positioning system-based automatic vehicle location data. *Journal of Intelligent Transportation Systems, 11*(4):157—167.

Patnaik, J., Chein, S., & Bladihas, A. (2004).Estimation of bus arrival times using APC data. *Journal of Public Transportation*, *7*, 1, 1–20. 7.

Wall, Z., & Dailey, D.J. (1999). An algorithm for predicting the arrival time of mass transit vehicles using automatic vehicle location data. *78th Annual Meeting of the Transportation Research Board, National Research Council.*

Shalaby, A., & Farhan, A. (2004). Bus travel time prediction for dynamic operations control and passenger information systems. *82nd Annual Meeting of the Transportation Research Board, National Research Council*.

Chien, S.I.J. & Kuchipudi, C.M. (2003). Dynamic travel time prediction with real-time and historic data. *Journal of Transportation Engineering, 129,* 6, 608-616.

Vanajakshi, L.,Subramanian, S.C, & Sivanandan, R. (2009). Travel time prediction under heterogeneous traffic conditions using global positioning system data from buses. *IET Intelligent Transport Systems* 3, 1 -9.

Chen, M., Liu, X., & Xia, J . (2004). A dynamic bus arrival time prediction model based on APC data. *Computer Aided Civil and Infrastructure Engineering, 19*,364 376.

Kwon, J., Coifman, B. & Bickel, P. (2000). Day to day travel time trends and travel time prediction from loop detector data. *Transportation Research Record,* 1717,120-129.

Frechette, L. & A. Khan, A.(1998). Bayesian regression based urban traffic models. *Transportation Research Record, 1644,* 157-165.

Park,T., Lee, S. & Moon, Y. (2004). Real time estimation of bus arrival time under mobile environment. *In Computational Science and Its Applications, International Conference, vol. 3043,* 1088–1096.